

# Saahil Mehta

📍 Reading, England, United Kingdom ✉ saahil.mehta8520@gmail.com ☎ +44 07537830191 🌐 in/saahil-mehta 📄 saahil.framer.wiki

## SUMMARY

Lead/Staff Data & Platform Engineer (7+ years) specialising in building resilient, observable and governed data & AI platforms for analytics and digital-first organisations. Deep experience across Google Cloud Platform (GCP) and Azure, Terraform, Python, SQL and dbt (all flavours), with a track record of greenfield builds, complex migrations, infrastructure-as-code, and SRE-grade reliability and cost control. Trusted to own architecture end-to-end, define and enforce standards, and mentor engineers. Actively delivering production-grade LLM/RAG and agentic systems (including MCP-based services, on-device/OCR pipelines, and RAG-backed knowledge platforms), applying rigorous data engineering, security and compliance principles to modern AI workloads.

## EXPERIENCE

### South Pole Group

London

#### Lead Specialist, Data Engineering

March 2024 - Present

- Own the core data platform and marts used by Finance, Carbon Trading, Projects, Sales and Advisory; set architecture, standards, and reliability targets.
- Migrated 18k+ resources from Pulumi to modular Terraform with GitHub Actions CI/CD, standardising IAM, networking, GKE/Cloud Run, logging, drift detection and policy-as-code.
- Delivered LLM/RAG microservices on GCP (FastAPI, FastMCP, Cloud Run) integrated into existing pipelines with full observability, enabling safe, low-latency AI services on production data.
- Implemented GDPR- and ISO-aligned governance: Terraform-based RBAC, row-level security and domain-isolated access for sensitive BigQuery datasets; contributed to AI governance readiness.
- Integrated Azure ERP and other enterprise systems into BigQuery via reusable cross-cloud patterns, keeping designs portable and compliant.
- Optimised Airflow + Dataproc + dbt stack, cutting key workloads by ~50% and improving stability for analytics and reporting.
- Built governed ELT from APIs, ecommerce, spreadsheets, blockchain and legacy systems into BigQuery with SLAs and data quality checks.
- Introduced FinOps and SRE-style monitoring: spend dashboards, alerts, anomaly detection, SLIs/SLOs and unified Grafana/Looker views.
- Mentored engineers and codified best practices for performance, security, modularity and observability across the platform.

### Publicis Groupe, Digitas UK

UK, England, London

#### Senior Data Engineer II

October 2022 - March 2024

- Built and optimised GCP + Python + Airflow pipelines, cutting processing time from 4 hours to 2 and consistently hitting SLAs for high-volume marketing analytics.
- Designed GitHub Actions CI/CD for data pipelines and Terraform, with automated checks, review bots and access controls, standardising secure delivery across teams.
- Consolidated cross-channel campaign data (search, social, programmatic, web, CRM) for clients incl. Samsung, Formula 1, Marriott, enabling reliable performance reporting and deeper optimisation.
- Automated user credential and access management, saving ~14 hours/month and tightening security/compliance.
- Developed reusable data scraping frameworks, increasing collection efficiency by ~40% and keeping pace with shifting stakeholder and media requirements.
- Ran training and produced playbooks for analysts and engineers, levelling up SQL, pipeline hygiene and analytics engineering standards across the team.

### Publicis Groupe, Digitas Mumbai

IN, MAHARASHTRA, Mumbai

#### Senior Associate Data Analytics Engineer

October 2021 - September 2022

- Cut cloud costs by 98.23% for a key client (~£500 → ~£7/month) through optimisation of storage, queries and scheduling, freeing budget for media and product experiments.
- Led a team of 3 to build automated pipelines and dashboards, enabling near real-time campaign analytics for regional stakeholders.
- Built a Python-based reporting bridge that eliminated repetitive manual reporting, materially improving delivery speed and consistency.
- Scaled Facebook Ads → Google Analytics load jobs from 3/week to 100+ /week, reducing deployment from 2 hours to under 10 minutes via better orchestration and packaging.

### WPP, GroupM, Mindshare

IN, MAHARASHTRA, Mumbai

#### Manager - Data & Technology

July 2021 - October 2021

- Built a GCP-based data lake for **Disney+ Hotstar**, reducing end-to-end processing from **2 days to 2 hours** and enabling timely content and campaign performance insights.
- Engineered robust ingestion and modelling of cross-platform data (app analytics, ad platforms, CRM), improving reliability and depth of reporting for regional stakeholders.
- Delivered enhanced insight and optimisation workflows that materially improved client retention and unlock upsell opportunities across key accounts.

Publicis Groupe, Digitas Mumbai	IN, MAHARASHTRA, Mumbai
Senior Associate Data Science	August 2019 - July 2021
<ul style="list-style-type: none"> <li>Automated data collection and transformation workflows, turning months of manual effort into minutes and significantly reducing error rates.</li> <li>Built and deployed multilingual sentiment analysis pipelines to track brand and campaign perception, roughly doubling usable signal quality for planners.</li> <li>Created scalable Power BI dashboards adopted across teams, cutting manual update time by ~50% and standardising performance reporting.</li> </ul>	

PROJECT

AI-Native Knowledge & RAG System (in-progress)	
Personal / OSS-aligned • <a href="#">knowsee.framer.ai</a>	
<ul style="list-style-type: none"> <li>Building an internal knowledge platform using agent frameworks, embeddings and vector stores to unify docs, standards and tribal knowledge into a secure, searchable system.</li> <li>Implemented RAG pipelines with careful chunking, metadata design and evaluation, respecting licensing and data residency constraints while enabling high-signal responses for engineering and business users.</li> <li>Integrated with open-source agent starter kits (<a href="#">github.com/GoogleCloudPlatform/contributions</a>) and MCP-style patterns; contributed improvements upstream, strengthening reliability, observability and developer experience for agentic workloads.</li> </ul>	
MERL Visitor Insight Engine	June 2025 - Present
Museum of English Rural Life, Reading (Volunteer)	
<ul style="list-style-type: none"> <li>Designed an on-device + cloud workflow using Apple Vision ML OCR and a two-stage multimodal LLM validation layer to convert handwritten visitor comment cards into clean, structured datasets.</li> <li>Mapped outputs to the Generic Learning Outcomes framework to quantify engagement, sentiment and learning impact across galleries, informing exhibit and education programme improvements.</li> <li>Introduced a guided workflow for a neurodivergent (autistic) volunteer, breaking down tasks into accessible steps and increasing their confidence with basic data tooling and processes.</li> </ul>	

EDUCATION

Master of Science in Data Analytics	UK, Glasgow
University of Strathclyde • First-Class Honours/ Distinction	2023
Bachelor of Computer Engineering	IN, MAHARASHTRA, Mumbai
University of Mumbai • 3.25	2019
Three-year Diploma in Computer Engineering	IN, MAHARASHTRA, Mumbai
SBM Polytechnic • 7.77	2016

SKILLS

Cloud & Infrastructure
Google Cloud Platform (BigQuery, IAM, Cloud Run, Cloud Functions, GKE, Dataproc, Composer, Cloud Build, Artifact Registry, Monitoring); Azure (selected services); Docker & Kubernetes; CI/CD with GitHub Actions.
Data Engineering & Analytics
Python, SQL, dbt, Apache Airflow; ELT/ETL design; data modelling (marts, warehouse layers, partitioning); batch & streaming (Pub/Sub, Spark); API integration; data quality and testing; performance and cost optimisation.
AI & LLM Systems
LLMs & SLMs, retrieval-augmented generation (RAG), embeddings & vector stores, agentic patterns (MCP, agent starter kits), FastAPI-based microservices, secure integration of AI services with production data, model evaluation and guardrail design.
Governance, Reliability & FinOps
RBAC & IAM, encryption, GDPR-first design, ISO 27001-aligned controls, exposure to ISO 42001 (AI governance), SLAs/SLIs/SLOs, policy-as-code, cost monitoring and optimisation, auditable data and AI pipelines.
Observability & Monitoring
GCP Logging & Monitoring, MQL, Prometheus, Grafana, structured logging, anomaly detection for data, infra and AI workloads.
Analytics & BI
BigQuery ecosystem, Looker Studio, Power BI, Tableau; Google Analytics & Adobe Analytics for marketing and product analytics, Alteryx.
Leadership & Collaboration
Mentoring engineers, leading cross-functional delivery, setting technical standards, running trainings, working closely with stakeholders in marketing, product, finance and sustainability teams.